INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Confidence intervals for adaptive regression estimation

Anatoli Juditsky , Oleg Lepski

No 3580

Décembre 1998

THÈME 4

apport
de recherche

ISSN 0249-6399

Confidence intervals for adaptive regression estimation

Anatoli Juditsky , Oleg Lepski*

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet is2

Rapport de recherche n° 3580 — Décembre 1998 — 29 pages

Abstract: The problem of adaptive estimation of the regression function f from noisy observation is concerned. A wavelet adaptive estimator of unknown function with the confidence interval for the L_2-error are provided. We show that if f belongs to a Sobolev class, the proposed estimate and the associated confidence interval are minimax.

Key-words: Adaptive estimation, nonparametric regression, confidence intervals, wavelet estimators

(Résumé : tsvp)

* Humboldt Universität zu Berlin, Spandauer str. 1, 10178 Berlin, Germany

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Les intervalles de confiance pour l'estimation adaptative de régression

Résumé : Nous considérons un problème d'estimation adaptative d'une fonction de régression f à partir d'observations bruitées. Nous proposons un estimateur adaptatif par ondelettes d'une la fonction inconnue avec un intervalle de confiance associé pour la norme L_2 de l'erreur d'estimation. On démontre que cet estimateur et son intervalle de confiance sont minimax si f appartient à une classe de Sobolev.

Mots-clé : Estimation adaptative, regréssion non paramétrique, intervalles de confiance, estimateurs par ondelettes

1 Introduction

We consider the problem of recovering of unknown function $f(x) : [0,1] \to \mathbf{R}$ from noisy observations

$$y_i = f(\frac{i}{N}) + w_i, \quad i = 1, ..., N, \tag{1}$$

where $(w_i, i = 1, ..., N$ is the vector of independent and identically distributed Gaussian random variables with $Ew_1 = 0$ and $Ew_1^2 = \sigma_w^2$.

The basic problem which has been extensively studied in the literature on nonparametric estimation is to provide an estimate $\widehat{f}_N$ of f given the observations (y_i). In particular, when the problem of minimax estimation is concerned, a usual approach is to introduce a functional class $\mathcal{F}$, defined with a set of parameters $\mathcal{U} = \{u_i\}$, then to design the minimax estimation algorithms on this class which depend explicitly on the constants u_i. When the accuracy (or rather inaccuracy) of estimation is measured by the L_2-norm of the error $\widehat{f}_N - f$, the optimal solution of this problem is available for a variety of functional classes (cf., for instance, [9] and [5])

To be more precise, suppose that f belongs to the Sobolev class $\mathcal{F}(s, L)$ (refer to Section 2 for definitions), defined with the parameters s (the regularity) and L (the Sobolev constant). We consider the risk

$$\rho(\widehat{f}_N, f) = E_f \|\widehat{f}_N - f\|^2,$$

where $\|\cdot\|$ is a norm or a quasi-norm. It is well known how to construct the minimax on the class $\mathcal{F}(s, L)$ estimator $\widehat{f}_N$, i.e the function $\widehat{f}_N^*$ which is the minimizer of

$$R(\widehat{f}_N, \mathcal{F}) = \sup_{f \in \mathcal{F}} E_f \rho(\widehat{f}_N, f)$$

for a large variety of risks $\rho(\cdot)$. This estimator attains the rate of convergence $R(\widehat{f}_N^*, \mathcal{F}) = O(N^{-\frac{s}{2s+1}})$. However, the class $\mathcal{F}(s, L)$ contains also functions which can be estimated with better rate. Indeed, this class contains, for instance, the Sobolev balls $\mathcal{F}(s', L)$ with $s' > s$. Then if the information that the unknown function belongs to such an embedded class was available, one can expect to find an estimate $\widehat{f}_N$ which attain the minimax rate of convergence which corresponds to the parameters (s', L) of this smaller class.

Then the following questions arise:

1. How to design an "adaptive" estimation algorithm which only uses the observations and which deliver an estimate $\widehat{f}_N$ of not worse quality than the parameter-dependent estimate, which uses the knowledge of parameters (s, L) which describe the Sobolev class.

2. If such an estimate $\widehat{f}_N$ is given, how to access its accuracy $m_N = \|\widehat{f}_N - f\|$ (here $\|\cdot\|$ stands for some norm or some quasi-norm of the error).

In the framework defined above the answer to the first question is given, for instance, in [13], [6], [12] and [1]. In those papers a variety of estimates $\widehat{f}_N$ are proposed, such that the ratio of the estimate risk $R(\widehat{f}_N, \mathcal{F})$ and the minimax risk $R(\widehat{f}^*_N, \mathcal{F})$ remains finite as $N \to \infty$. Following [13] we call such estimates *adaptive in order*. Note, however, that those adaptive estimation algorithms do not typically provide any information about the error m_N.

If it is known *a priori* that $f \in \mathcal{F}(s, L)$ then one can take m^*_N as the minimax rate of convergence on the class $\mathcal{F}(s, L)$ which is $O(N^{-\frac{s}{2s+1}})$. On the other hand, as we have seen above, if f belongs to the class $\mathcal{F}(s', L)$ with higher regularity then the adaptive estimate $\widehat{f}_N$ will attain better rate of convergence and the bound m^*_N would be rather pessimistic. On the other hand, it is also known that if the accuracy of estimation is characterized with the L_∞-norm of the error the bound $O(N^{-\frac{s}{2s+1}})$ of the error m_N cannot be improved in the minimax sense (cf. [14]). This motivates our choice of the L_2-norm of the error $m_N = ||\widehat{f}_N - f||_2$ as the measure of the quality of the estimate to be accessed. In fact we want to point out a value $\tau_N(\alpha)$ (a confidence interval) such that for any $\alpha > 0$ the ball $B(\widehat{f}_N, \tau_N(\alpha))$ in L_2-space, centered at $\widehat{f}_N$ with radius $c(\alpha)\tau_N$, satisfies

$$P_f(f \in B(\widehat{f}_N, \tau_N(\alpha))) \geq 1 - \alpha. \tag{2}$$

In order to characterize the quality of the bound τ_N one can use the quadratic error:

$$r(\tau_N, f) = [E_f(\tau_N - m_N)^2]^{1/2}. \tag{3}$$

So in the present paper our objective is twofold: we are to design an adaptive in order estimate $\widehat{f}_N$ of f *and* to provide the method to construct confidence intervals $\tau_N(\alpha)$ for the estimator $\widehat{f}_N$ from observations $y_1, ..., y_N$. Of course, we aim to obtain the quantities τ_N which are "good" in the sense of (3). One can consider the following "common-sense" strategy to solve the problem above: given the sample (y_i), $i = 1, ..., N$, we split it into two independent subsamples, say $(y_i^{(1)})$ and $(y_i^{(2)})$. Then we use one of known adaptive estimates to retrieve the estimate $\widehat{f}_N$ of unknown function f from observations $(y_i^{(1)})$. Next we compute the estimate $\widehat{m}_N$ of m_N using the observations $(y_i^{(2)} - \widehat{f}_N(i/N))$ from the second subsample. It is known that if the unknown function f belongs to the Sobolev class $\mathcal{F}(s, L)$ then the adaptive estimate $\widehat{f}_N$ (which does not use the knowledge of s and L) satisfies $R(\widehat{f}_N, \mathcal{F}) = O(N^{-2s/(2s+1)})$. On the other hand, if θ_N is an adaptive estimate of the L_2-norm $||f||_2$ of f then (cf. [8])

$$\sup_{f \in \mathcal{F}} E_f(\theta_N - ||f||_2)^2 = O\left(\left(\frac{\sqrt{\log N}}{N}\right)^{\frac{4s}{4s+1}}\right). \tag{4}$$

On can expect to obtain the same bound (4) for the error $\widehat{m}_N - m_N$ in the problem above (in Section 5 we show how a simple adaptive estimator with an adaptive estimate of the L_2-error norm can be constructed on Sobolev classes without splitting the data).

Unfortunately, the estimate $\widehat{m}_N$ have an obvious drawback: the quantity $c(\alpha)\widehat{m}_N$ cannot provide an *upper* bound for the error m_N (at least for small $m_N = O\left(N^{-\frac{2s}{4s+1}}\right)$). Thus cannot be used in the construction of the confidence τ_N interval in (2).

In Theorem 1 below that a minimax on the class $\mathcal{F}(s,L)$ lower bound for the rate of convergence of an estimate $\widehat{m}_N$ of the error m_N:

$$\sup_{f\in\mathcal{F}} E_f(\widehat{m}_N - m_N)^2 \geq c_0\left(N^{-\frac{4s}{4s+1}}\right).$$

As a result, if no information on the parameters (s,L) of the functional classes available, the problem of construction of confidence intervals cannot be solved in the minimax sense. On the other hand, we show in Theorem 2 that if it is *a priori* known that $f \in \mathcal{F}(s^*, L^*)$, a confidence interval τ_N can be constructed such that

$$\sup_{f\in\mathcal{F}} E_f(\tau_N - m_N)^2 = O\left(N^{-\frac{4s^*}{4s^*+1}}\right).$$

And the lower bound in Theorem 1 shows that this rate of convergence cannot be essentially improved.

The paper is organized as follows: in Section 2 we recall some basic properties of adaptive wavelet estimators. Then in Section 3 a minimax lower bound for the rate of convergence of L_2-error estimators is established. Next in Section 4 we provide an adaptive estimator $\widehat{f}_N$ of f with an associated confidence interval τ_N for its L_2-error on the Sobolev class. Finally, in Section 5 we provide a simple adaptive estimator with an adaptive estimate $\widehat{m}_N$ of its L_2-error norm on the family of Sobolev classes.

2 Adaptive wavelet estimators

We start with the definition of functional classes used.

2.1 Decompositions of Sobolev classes

Let ϕ_k, ψ_{jk} be a system of compactly supported orthogonal wavelets (supp$\phi \subseteq [-A, A]$ and supp$\psi \subseteq [-A, A]$), i.e. $\phi_k(x)$ and $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, $j = 1, \ldots,$ constitute (inhomogeneous) orthonormal wavelet basis of $L_2(0,1)$ [15], [3]. Let $m = \max(1, s_{\max})$. We suppose that ϕ and $\psi \in C^m$. This implies (see Ch. 7, [3]) that $\psi(x)$ has $l = [s_{\max}]$ vanishing moments (here $[\cdot]$ is an integer part). We just note that wavelet basis on $[0,1]$ with such properties can be constructed (see, for instance, [2]). Since the regression function and the wavelets are compactly supported, there are at most $(2^j + 2A - 1)$ nonzero coefficients at each resolution level j of the wavelet expansion of f. We suppose with some stretch that this number is exactly 2^j, thus

$$f(x) = \alpha\phi(x) + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1}\beta_{jk}\psi_{jk}(x),$$

where

$$\alpha = \int f(x)\phi(x)dx, \quad \beta_{jk} = \int f(x)\psi_{jk}(x)dx.$$

From now on we suppose that the unknown function f belongs to some set $\mathcal{F} \in L_2(0,1)$ which is defined through the coefficients α and β_{jk} of the wavelet decomposition of f:

$$f(x) = \alpha\phi(x) + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x). \tag{5}$$

Note that due to the orthonormality of functions ψ_{jk} and ϕ,

$$||f||_2^2 = \alpha^2 + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}^2.$$

We suppose that $\mathcal{F}$ is a "Sobolev body"[1] of wavelet coefficients:

$$\begin{aligned} \mathcal{F} &= \mathcal{F}(s, L) \\ &= \{f \text{ such that } ||f||_{s,2} \le L\}, \text{ where } ||f||_{s,2}^2 = \alpha^2 + \sum_{j=0}^{\infty} 2^{2js} \sum_{k=0}^{2^j-1} \beta_{jk}^2. \end{aligned} \tag{6}$$

Note that if W^s, $s \ge -1/2$, is the Sobolev space (see [16]), then there is $C > 0$ such that

$$||f||_{W^s} \ge C||f||_{s,2}, \tag{7}$$

where $||f||_{W^s}$ is the norm of the Sobolev space (cf. Theorem 2 in [4]. See also [7] for a discussion and useful references). In fact, the norms for a wide variety of functional spaces can be "efficiently" expressed in terms of the coefficients of wavelet decompositions [15]. For instance, $H(s, L)$ is a Hölder class of functions (cf. [11]), then $H(s, L) \subseteq \mathcal{F}'(s, c_0 L)$, where

$$\mathcal{F}'(s, L) = \{\beta : \ |\alpha| + \max_{j\ge 0} 2^{j(s+1/2)}||\beta_{j\cdot}||_\infty \le L\} \tag{8}$$

and c_0 is a constant which depends only on the particular choice of wavelet ψ. In what follows with some abuse of notations we refer to $\mathcal{F}(s, L)$ as the Hölder class.

2.2 Wavelet estimators

Consider the following problem: given the observations (1) to design an estimate $\widehat{f}_N$ of f which uses only the observations $y_1, ..., y_N$ (but not the knowledge of the parameters s and L of the class), such that for any class $\mathcal{F}(s, L)$ the ratio of the estimate risk

$$R(\widehat{f}_N, \mathcal{F}) = \sup_{f\in\mathcal{F}} E_f||\widehat{f}_N - f||_2^2$$

[1] we borrow the terminology of D. Donoho and I. Johnstone [5].

to the minimax risk

$$R(\mathcal{F}) = \inf_{\widehat{f}_N} R(\widehat{f}_N, \mathcal{F})$$

remains finite as $N \to \infty$. Following [13] we call such estimates *adaptive in order*.

In the above problem the minimax rates of convergence were established in [13]. These rates are attained, for instance, by adaptive wavelet estimates $\widehat{f}_N$, designed in [6], [12] and [1]. For our model these estimates are constructed as follows: first we compute the coefficients

$$\widehat{\alpha}_k = N^{-1} \sum_{i=1}^{N} y_i \phi_k(\frac{i}{N}), \quad y_{jk} = N^{-1} \sum_{i=1}^{N} y_i \psi_{jk}(\frac{i}{N}), \text{ for } j = 0, \ldots, j_0,$$

where j_0 is such that $\frac{N}{2} < 2^{j_0} \le N$. Then y_{jk} are shrinked to zero using the thresholding rule:

$$\widehat{\beta}_{jk} = \delta(y_{jk}, \lambda_j). \tag{9}$$

Here $\delta(\cdot)$ can be hard- or soft-thresholding rule, respectively,

$$\delta(x, \lambda) = x 1_{|x| \ge \lambda} \quad \text{or} \quad \delta(x, \lambda) = \operatorname{sign}(x)(x - \lambda)_+.$$

The threshold λ_j is selected from the observations using a kind of cross-validation procedure which is different for the estimates provided in the papers cited above. Finally one put

$$\widehat{f}_N(x) = \sum_k \widehat{\alpha}_k \phi_k(x) + \sum_{j=0}^{j_0} \sum_k \widehat{\beta}_{jk} \psi_{jk}(x). \tag{10}$$

The risk $R(\cdot)$ of these estimates on the Sobolev class satisfy

$$R(\widehat{f}_N, \mathcal{F}) \le C L^{2/(2s+1)} \left(\frac{\sigma_w^2}{N} \right)^{\frac{2s}{2s+1}} + O\left(\frac{\sigma_w^2 \log^2 N}{N} \right).$$

Another adaptive estimate which attains the same rate of convergence on $\mathcal{F}(s, L)$ (cf. [12]) can be obtained if instead of the estimate (9) of wavelet coefficients β_{jk} we use

$$\widehat{\beta}_{jk} = y_{jk} 1_{\rho_j \ge 2^j \sigma_w^2 / N},$$

where

$$\rho_j = \sum_k (y_{jk}^2 - \frac{\sigma_w^2}{N}).$$

In other words, the same thresholding rule is applied to *all* coefficients at the level j. In what follows we consider different estimates of the "inaccuracy" $m_N = ||\widehat{f}_N - f||_2$ of this estimator.

3 Lower bound for confidence interval estimation

Suppose that the observations $y_i = f(\frac{i}{N}) + w_i$, $i = 1, ..., N$ of the function f are available. It is known *a priori* that $f \in \mathcal{F}(s, L)$.

Let $\widehat{f}_N$ be an estimate of f. Our objective here to establish the lower bound on the rate of convergence of the estimate of the error $||\widehat{f}_N - f||_2$. This bound cannot be given for all estimates $\widehat{f}_N$; indeed, the error of a trivial estimate $\widehat{f}_N(i/N) = y_i$ can be estimated with parametric rate. However, if we limit our consideration to a class of "not-trivial" estimates, which, of course, are the only estimates being of interest, such a bound can be established. Such a class of "reasonably good" estimates can be defined in many ways. We consider here the following

Assumption 1. The estimate $\widehat{f}_N$ is "almost minimax" on $\mathcal{F}(s, L)$, i.e. if

$$\nu_N = L^{1/(2s+1)} \left(\frac{\sigma_w^2 \log N}{N} \right)^{s/(2s+1)},$$

then for some $C < \infty$

$$\sup_{f \in \mathcal{F}(s,L)} \nu_N^{-1} [E_f ||\widehat{f} - f||_2^2]^{1/2} \le C.$$

Note that this assumption holds for known adaptive estimates (cf. the estimates proposed in [7], [6] or [12]).

Let now for some $0 < \delta < 1$ $f_0 \in \mathcal{F}(s, (1-\delta)L)$. We say that f belongs to $\delta\mathcal{F}_{f_0}(s, L)$ if $f - f_0 \in \mathcal{F}(s, \delta L)$.

Theorem 1 *Suppose that Assumption 1 holds for the estimate $\widehat{f}_N$ of f. Then there is an absolute constant c_0 such that for any $0 < \delta < 1$, $f_0 \in \mathcal{F}(s, (1-\delta)L)$ and any estimate $\widehat{m}_N$ of $m_N = ||f - \widehat{f}_N||_2$ it holds*

$$\sup_{f \in \delta\mathcal{F}_{f_0}(s,L)} \left[E_f (m_N - \widehat{m}_N)^2 \right]^{1/2} \ge \begin{cases} c_0 (L\delta)^{1/(4s+1)} \left(\frac{\sigma_w^2}{N} \right)^{2s/(4s+1)}, & \textit{for} \quad s \ge 1/4 \\ c_0 \delta L N^{-s} & \textit{for} \quad s < 1/4. \end{cases} \tag{11}$$

The proof of the theorems are put in Section 6.

4 Adaptive estimator with a confidence interval

We present in this section an adaptive algorithm to estimate unknown function f and an estimate $\widehat{m}_N$ (and an upper estimate τ_N) of the estimation error m_N from observations (y_i), $i = 1, ..., N$ as in (1). We suppose that it is known *a priori* that $f \in F(s^*, L^*)$.

Algorithm 1 Put $\sigma^2 = \frac{\sigma_w^2}{N}$, take

$$j_0 \text{ such that } \frac{N}{2} < 2^{j_0} \leq N \tag{12}$$

and

$$j^* \text{ such that } \left(\frac{N(L^*)^2}{\sigma_w^2}\right)^{2/(4s^*+1)} \leq 2^{j^*} < 2\left(\frac{N(L^*)^2}{\sigma_w^2}\right)^{2/(4s^*+1)}, \tag{13}$$

if $j^* > j_0$ set $j^* = j_0$.

1. Compute the empirical wavelet coefficients

$$\widehat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} y_i \phi(\frac{i}{N}) \text{ and } y_{jk} = \frac{1}{N}\sum_{i=1}^{N} y_i \psi_{jk}(\frac{i}{N}), \quad 0 \leq k \leq 2^j - 1, \quad j = 0, ..., j_0. \tag{14}$$

2. For $j = 0, ..., j_0$ compute

$$\rho_j = \sum_{k=0}^{2^j-1} y_{jk}^2 - \sigma^2 = \|y_{j\cdot}\|_2^2 - 2^j\sigma^2 \tag{15}$$

and the estimates of wavelet coefficients

$$\widehat{\beta}_{jk} = y_{jk} 1_{\rho_j \geq 2^j\sigma^2}. \tag{16}$$

3. To terminate set

$$\widehat{f}_N(x) = \widehat{\alpha}\phi(x) + \sum_{j \leq j_0}\sum_k \widehat{\beta}_{jk}\psi_{jk}(x) \tag{17}$$

and

$$\widehat{m}_N^2 = \left[\sum_{j=0}^{j_0} 2^j\sigma^2 1_{\rho_j \geq 2^j\sigma^2} + \sum_{j=0}^{j^*} \rho_j 1_{\rho_j < 2^j\sigma^2}\right]_+. \tag{18}$$

Theorem 2 *Let $\mathcal{F}(s,L)$ be a Sobolev class such that $\mathcal{F}(s,L) \subseteq \mathcal{F}(s^*,L^*)$. Then*

$$\sup_{f \in \mathcal{F}(s,L)} [E_f\|f - \widehat{f}_N\|_2^2]^{1/2} \leq CL^{1/(2s+1)}\left(\frac{\sigma_w^2}{N}\right)^{s/(2s+1)} + \epsilon(N), \tag{19}$$

where $|\epsilon(N)| \leq \frac{C_0\sigma_w \log N}{\sqrt{N}}$. Furthermore, the estimate $\widehat{m}_N$ satisfies:

$$\sup_{f \in \mathcal{F}(s^*,L^*)} \left[E_f(\widehat{m}_N - m_N)^2\right]^{1/2} \leq C_1\left((L^*)^{1/(4s^*+1)}\left(\frac{\sigma_w^2}{N}\right)^{2s^*/(4s^*+1)} + L^*N^{-s^*}\right) + \epsilon(N). \tag{20}$$

Here C_0 and C_1 are constants which do not depend on N, L^* and σ_w and can be computed explicitly for a given value of s^* and wavelet ψ.

Remark: Using the bound (20) for the error $\widehat{m}_N - m_N$ we can modify the estimation algorithm above to construct a confidence interval τ_N. Indeed, if we set

$$\tau_N(\alpha) = \widehat{m}_N + \sqrt{\alpha}\left[C_1(L^*)^{1/(4s^*+1)}\left(\frac{\sigma_w^2}{N}\right)^{2s^*/(4s^*+1)} + C_1 L^* N^{-s^*} + \epsilon(N)\right], \qquad (21)$$

we obtain the following evident

Corollary 1 *The quantity $\tau_N(\alpha)$, delivered by Algorithm 1 and (21), satisfies*

$$\left[E(\tau_N - m_N)^2\right]^{1/2} \leq (1+\sqrt{\alpha})\left[C_1(L^*)^{1/(4s^*+1)}\left(\frac{\sigma_w^2}{N}\right)^{2s^*/(4s^*+1)} + C_1 L^* N^{-s^*} + \epsilon(N)\right]$$

and for $\alpha > 1$

$$P(m_N > \tau_N) \leq \frac{1}{\alpha}.$$

Remark: note that the upper bound for the estimation error of the adaptive estimator $\widehat{f}_N$, established in Theorem 2 is tight. Furthermore, we conclude from the lower bound of Theorem 1 the estimate $\widehat{m}_N$, provided by Algorithm 1 is minimax optimal (20) (up to a constant).

Remark that for $s^* > 1/4$ the main error term in (20) is equivalent to the function estimation error if we substitute $2s^*$ for s in the exponent (cf. the first term in the right-hand side of (19)).

One can see from (20) (cf. the lower bound (11)) that the behaviour of the adaptive estimate $\widehat{m}_N$ changes dramatically as the quantity

$$(\sigma_w L)^4 N^{4s^*-1}$$

becomes $O(1)$. If L and σ_w are $O(1)$ this can happen when $s \approx 1/4$. For $s^* > 1/4$ the first term in the right-hand side of (24) is dominant, for smaller values of s^* the second term LN^{-s^*} become major. For these values of s^* the estimate “sinks into the grid", in other words, the bias term of the error $\widehat{m}_N - m_N$ which is due to the approximation of the function f from its values on the grid (i/N) is larger than any other error components. The situation changes dramatically if our objective to recover the values $f(i/N)$ of the function **on the grid**. In this case one can easily deduce from the proof of Theorem 2 that for $s^* < 1/4$

$$\left[E_f(\widehat{m}_N - m_N)^2\right]^{1/2} = O\left(\left(\frac{\sigma_w^2\sqrt{\log N}}{N}\right)^{1/4}\right).$$

5 Simple adaptive estimate

Adaptive estimation algorithm. We provide here another adaptive estimator f with the estimation $\widehat{m}_N$ of the L_2-error. When compared to Algorithm 1 in section 4, the estimate $\widehat{m}_N$, computed by this method is adaptive with respect to the parameters s, L of the Sobolev class. Note however, that this estimate cannot be used to compute a confidence interval for the adaptive estimate of $\widehat{f}_N$.

Algorithm 2 `Choose` $s_{\max} < \infty$ `and a wavelet` $\psi(x)$ `with` $l = [s_{\max}]$ `vanishing moments. Put` $\sigma^2 = \frac{\sigma_w^2}{N}$, $\lambda = \kappa\sqrt{\log N}$ `for` $\kappa > 16$ `and take` j_0 `as in (12), i.e.` $\frac{N}{2} < 2^{j_0} \leq N$

1. `Compute the empirical wavelet coefficients`

$$\widehat{\alpha} = \frac{1}{N}\sum_{i=1}^{N} y_i \phi(\frac{i}{N}) \quad \text{and} \quad y_{jk} = \frac{1}{N}\sum_{i=1}^{N} y_i \psi_{jk}(\frac{i}{N}), \quad 0 \leq k \leq 2^j - 1, \quad j = 0, ..., j_0.$$

2. `For` $j = 0, ..., j_0$ `compute`

$$\rho_j = \sum_{k=0}^{2^j-1} y_{jk}^2 - \sigma^2 = \|y_{j\cdot}^2\|_2^2 - 2^j\sigma^2$$

`and the estimates of wavelet coefficients`

$$\widehat{\beta}_{jk} = y_{jk} 1_{\rho_j \geq 2^j\sigma^2}.$$

3. `To terminate set`

$$\widehat{f}_N(x) = \widehat{\alpha}\phi(x) + \sum_{j \leq j_0}\sum_k \widehat{\beta}_{jk}\psi_{jk}(x)$$

`and`

$$\widehat{m}_N^2 = \left[\sum_{j=0}^{j_0} 2^j\sigma^2 1_{\rho_j \geq 2^j\sigma^2} + \sum_{j=0}^{j_0} \rho_j 1_{\lambda 2^{j/2}\sigma^2 \leq \rho_j < 2^j\sigma^2}\right]_+ \tag{22}$$

We have the following

Theorem 3 *Let $\mathcal{F}(s, L)$, $s < s_{\max}$ be a Sobolev class. The adaptive estimate $\widehat{f}_N$ satisfies:*

$$\sup_{f \in \mathcal{F}(s,L)} [E_f\|f - \widehat{f}_N\|_2^2]^{1/2} \leq CL^{1/(2s+1)}\left(\frac{\sigma_w^2}{N}\right)^{s/(2s+1)} + \epsilon(N), \tag{23}$$

where $\epsilon(N) = O\left(\frac{\sigma_w \log N}{\sqrt{N}}\right)$. Furthermore,

$$\sup_{f \in \mathcal{F}(s,L)} \left[E_f(\widehat{m}_N - m_N)^2\right]^{1/2} \leq C\left(L^{1/(4s+1)}\left(\frac{\sigma_w^2\sqrt{\log N}}{N}\right)^{2s/(4s+1)} + LN^{-s}\right) + \epsilon(N), \tag{24}$$

Comments: The lower bound for the adaptive estimation of the L_2-norm of a function, obtained in [8] suggests that the estimator $\widehat{m}_N$ above is adaptive in order.

The estimate $\widehat{m}_N$ of m_N cannot be considered as a confidence interval for the estimate $\widehat{f}_N$. Indeed, one can easily see that due to the thresholding procedure in (22), for certain functions $f \in \mathcal{F}$, $\widehat{m}_N = 0$ with positive probability, while $m_N = ||\widehat{f}_N - f||_2$ is strictly positive.

6 Proof of Theorems

In what follows C, C', C'' stand for positive constants which values may depend only on the parameters s, p and q of Besov classes.

6.1 Proof of Theorem 1

The proof of the lower bound (11) in the case $s < 1/4$ is evident. Indeed, a function $\delta f \in \delta\mathcal{F}_{f_0}(s, L)$ which vanishes on the grid $\frac{i}{N}$, $i = 1, ..., N$ with the norm $||\delta f||_2 \geq c_0 \delta L N^{-s}$ can always be constructed. However, the functions f_0 and $f_0 + \delta f$ cannot be distinguished from the observations (y_i).

To show the bound in the case $s \geq 1/4$ we transfer the problem in the space of wavelet coefficients.

With some abuse of notations we say that $\beta \in \mathbf{R}^N$ belongs to $\mathcal{F}(s, L)$ if

$$\max_{j \geq 0} 2^{j(s+1/2)} ||\beta_{j\cdot}||_\infty \leq L.$$

In fact, this implies that $f(x) = \sum_j \sum_k \beta_{jk} \psi_{jk}(x)$ belongs to the Hölder class $H(s, c_0 L)$ (cf. (8)) with some constant c_0 which depends on the choice of ψ. Let $\beta^0 \in \mathbf{R}^N$ be a vector of wavelet coefficients. We say that $\beta \in \delta\mathcal{F}_{\beta^0}(s, L)$ if for some $\delta > 0$ $\beta - \beta^0 \in \mathcal{F}(s, \delta L)$. Now suppose that for some $0 < \delta \leq 1$ $\beta_0 \in \mathbf{R}^N$ belongs to $\mathcal{F}(s, (1-\delta)L)$. Let $y \in \mathbf{R}^N = (y_{jk})$,

$$y_{jk} = \beta_{jk} + \sigma \zeta_{jk} \tag{25}$$

be the observation of the vector $\beta = (\beta_{jk}) \in \mathbf{R}^N$, $\beta \in \delta\mathcal{F}_{\beta^0}(s, L)$ (note that this also implies that $\beta \in \mathcal{F}(s, L)$). $\zeta = (\zeta_{jk})$ in (25) is a vector of independent and identically distributed Gaussian random variables, $E\zeta_1 = 0$, $E\zeta_1^2 = 1$.

Let θ_N an estimate of the quantity $||\hat{\beta} - \beta||_2$. The proof of Theorem 1 results from the following

Proposition 1 *For any $0 < \delta \leq 1$ and any $\beta^0 \in \mathcal{F}(s, (1-\delta)L)$, there are c_0, $c_1 > 0$ such that for all N sufficiently large, any estimate θ_N and any estimate $\widehat{\beta}$*

$$\sup_{\beta \in \delta\mathcal{F}_{\beta^0}(s,L)} E_\beta (\theta_N - ||\hat{\beta} - \beta||_2)^2 \geq c_0 \frac{\rho_N^2 \left(\rho_N^2 - c_1 \sigma (E_{\beta^0} ||\hat{\beta} - \beta^0||_2^2)^{1/2} \right)}{(\rho_N + (E_{\beta^0} ||\hat{\beta} - \beta^0||_2^2)^{1/2})^2},$$

where

$$\rho_N = \min\left\{(L\delta)^{1/(4s+1)}\sigma^{4s/(4s+1)}, \sigma N^{1/4}\right\}. \tag{26}$$

Let us show that the statement of Theorem 1 indeed follows from Proposition 1. We set

$$f_0(x) = \sum_{j=0}^{j_0} \beta^0_{jk}\psi_{jk}(x), \quad \text{and} \quad f(x) = \sum_{j=0}^{j_0} \beta_{jk}\psi_{jk}(x),$$

Then Assumption 1 implies that $\sigma\left(E_{\beta^0}\|\hat{\beta} - \beta^0\|_2^2\right)^{1/2} = o(\rho_N)$, and

$$\sup_{f\in\delta\mathcal{F}_{f_0}(s,L)} \left[E_f(\theta_N - \|\hat{f}_N - f\|_2)^2\right]^{1/2} \geq c_0\rho_N.$$

∎

6.2 Proof of Proposition 1

Let $\beta^0 = (\beta^0_{jk}) \in \mathbf{R}^N$, $\beta^0 \in \mathcal{F}(s, L(1-\delta))$ and j_0 satisfy

$$(L\delta)^{\frac{4}{4s+1}}\sigma^{-4/(4s+1)} \leq 2^{j^*} < 2(L\delta)^{\frac{4}{4s+1}}\sigma^{-4/(4s+1)}.$$

If $2^{j^*} > N$ we put $2^{j^*} = N$. We define

$$\tilde{\beta} = L\delta 2^{-j^*(s+1/2)}. \tag{27}$$

Next we set

$$\xi_{jk} = \begin{cases} 0, & \text{if} \quad j \neq j^* \\ \xi_k, & \text{if} \quad j = j^*, \end{cases}$$

where (ξ_k), $k = 0, ..., 2^{j^*}-1$ is a sequence of independent and identically distributed Bernoulli random variables, $P(\xi_0 = 1) = P(\xi_0 = -1) = 1/2$. Finally, we define the vector $\beta^{(\xi)}$ in the following way: make another independent drawing such that

$$\beta^{(\xi)} = \begin{cases} \beta_0 + \tilde{\beta}\xi & \text{with probability} \quad \frac{1}{2} \\ \beta_0 & \text{with probability} \quad \frac{1}{2} \end{cases}$$

Note that due to the definition (27) of $\tilde{\beta}$, the vector $\beta^{(\xi)}$ belongs to $\mathcal{F}(s, L)$. Consider the observation $y = (y_{jk})$ of $\beta^{(\xi)}$,

$$y_{jk} = \beta^{(\xi)}_{jk} + \sigma\zeta_{jk},$$

where (ζ_{jk}) is a sequence of independent and identically distributed Gaussian random variables (independent of ξ) with $E\zeta_{jk}=0$ and $E\zeta_{jk}^2=1$.

We can now write down the Bayesian risk of an estimate θ_N of $\|\hat\beta-\beta^{(\xi)}\|_2$

$$r_{\beta^0}(\delta,N)=\frac{1}{2}E_\xi\left\{E_{\beta^{(\xi)}}(\theta_N-\|\hat\beta-\beta^{(\xi)}\|_2)^2+E_{\beta^0}(\theta_N-\|\hat\beta-\beta^0\|_2)^2\right\},$$

where E_ξ stands for the expectation with respect to the distribution of ξ. Let us denote $\Delta_N=\theta_N-\|\hat\beta-\beta^{(\xi)}\|_2$, $\delta\beta=\hat\beta-\beta^0$ and let Z_ξ stand for the likelihood ratio

$$Z_\xi=\frac{dP_{\beta^{(\xi)}}}{dP_{\beta^0}}=\prod_{k=0}^{2^{j^*}-1}\exp\left(\frac{\zeta_{j^*k}\xi_k\tilde\beta}{\sigma}-\frac{\tilde\beta^2}{2\sigma^2}\right).$$

We define now the following events:

$$\begin{array}{rcl} A &=& \{\Omega:\ \frac{1}{2}\le E_\xi Z_\xi\le 4e\},\\ B &=& \{\Omega:\ \sum_{k=0}^{2^{j^*}-1}\zeta_{j^*k}^2\le 22^{j^*}\},\\ C &=& \{\Omega:\ \|\delta\beta\|_2\le 8\sqrt{e}(E_{\beta^0}\|\delta\beta\|_2^2)^{1/2}\}, \end{array} \qquad (28)$$

and $\Gamma=\{A\cap B\cap C\}$.

Lemma 1

$$r_{\beta^0}(\delta,N)\ge\frac{E_{\beta^0}\left\{E_\xi[Z_\xi(\rho_N^2-2\tilde\beta\delta\beta^T\xi)^2]1\{\Gamma\}\right\}}{(E_{\beta^0}\|\delta\beta\|_2^2)^{1/2}+\rho_N)^2}.$$

Proof: Since

$$\theta_N-\|\hat\beta-\beta^{(\xi)}\|_2=\theta_N-\|\hat\beta-\beta^0\|_2-(\|\hat\beta-[\beta^0+\tilde\beta\xi]\|_2-\|\hat\beta-\beta^0\|_2),$$

when changing the integration measure, we have for $r_{\beta^0}(\delta,N)$:

$$2r_{\beta^0}(\delta,N)=E_{\beta_0}\left\{E_\xi\left(Z_\xi(\Delta_N-\|\delta\beta-\tilde\beta\xi\|_2+\|\delta\beta\|_2)^2\right)+\Delta_N^2\right\}, \qquad (29)$$

where $\Delta_N=\theta_N-\|\widehat\beta-\beta^0\|_2$. Note that

$$\Delta_N^*=\frac{E_\xi\left[Z_\xi(\|\delta\beta-\tilde\beta\xi\|_2)-\|\delta\beta\|_2]\right)}{1+E_\xi Z_\xi}$$

is the minimizer of (29) with respect to Δ_N. When substituting Δ_N^* into (29) we get

$$\begin{aligned} 2r_{\beta^0}(\delta,N) &\ge E_{\beta^0}\left\{\frac{1}{1+E_\xi Z_\xi}E_\xi\left[Z_\xi(\|\delta\beta\|_2-\|\delta\beta-\tilde\beta\xi\|_2)^2\right]\right\}\\ &= E_{\beta^0}\left\{\frac{1}{1+E_\xi Z_\xi}E_\xi\left[Z_\xi\left(\frac{\|\delta\beta\|_2^2-\|\delta\beta-\tilde\beta\xi\|_2^2}{\|\delta\beta\|_2+\|\delta\beta-\tilde\beta\xi\|_2}\right)^2\right]\right\}. \qquad (30) \end{aligned}$$

Since the expression under the expectation E_{β^0} in (30) is positive, we can bound r_{β^0} from below as follows:

$$2r_{\beta^0}(\delta,N) \geq E_{\beta^0}\left\{\frac{1}{1+E_\xi Z_\xi}E_\xi\left[Z_\xi\left(\frac{||\delta\beta||_2^2-||\delta\beta-\tilde{\beta}\xi||_2^2}{||\delta\beta||_2+||\delta\beta-\tilde{\beta}\xi||_2}\right)^2\right]1\{\Gamma\}\right\}. \tag{31}$$

We use $||\tilde{\beta}\xi||_2^2=\rho_N^2$ and the bound for $||\delta\beta||_2$ on Γ to obtain from (31):

$$\begin{aligned} 2r_{\beta^0}(\delta,N) &\geq E_{\beta^0}\left\{\frac{1}{1+E_\xi Z_\xi}E_\xi\left[Z_\xi\left(\frac{\rho_N^2-2\tilde{\beta}\delta\beta^T\xi}{2||\delta\beta||_2+\rho_N}\right)^2\right]1\{\Gamma\}\right\} \\ &\geq \frac{E_{\beta_0}E_\xi\left\{Z_\xi(\rho_N^2-2\tilde{\beta}\delta\beta^T\xi)^2 1\{\Gamma\}\right\}}{(1+4e)(16\sqrt{e}(E_{\beta_0}||\delta\beta||_2^2)^{1/2}+\rho_N)^2}. \end{aligned} \tag{32}$$

■

Lemma 2 *Denote $I=E_\xi[\tilde{\beta}\delta\beta^T\xi Z_\xi]$. Then on Γ we have the following bound:*

$$I \leq 32\sqrt{2}e^{3/2}\sigma(E_{\beta^0}||\delta\beta||_2^2)^{1/2}.$$

Proof: Recall that by the definition of ξ

$$\tilde{\beta}\delta\beta^T\xi = \tilde{\beta}\sum_{k=0}^{2^{j^*}-1}\delta\beta_{j_0k}\xi_k,$$

so that, due to the independence of (ξ_k),

$$I = \tilde{\beta}\sum_{k=0}^{2^{j^*}-1}\delta\beta_{j^*k}\frac{\exp(\frac{\zeta_{j^*k}\tilde{\beta}}{\sigma}-\frac{\tilde{\beta}^2}{2\sigma^2})-\exp(-\frac{\zeta_{j^*k}\tilde{\beta}}{\sigma}-\frac{\tilde{\beta}^2}{2\sigma^2})}{2}\Pi_k,$$

where

$$\Pi_k = \prod_{l\neq k}^{2^{j^*}-1}\frac{\exp\left(\frac{\zeta_{j^*l}\tilde{\beta}}{\sigma}-\frac{\tilde{\beta}^2}{2\sigma^2}\right)+\exp\left(-\frac{\zeta_{j^*l}\tilde{\beta}}{\sigma}-\frac{\tilde{\beta}^2}{2\sigma^2}\right)}{2}.$$

Note that $|\mathrm{sh}(x)|\leq|x|\mathrm{ch}(x)$, thus

$$\left|\frac{\exp(\frac{\zeta_{j^*k}\tilde{\beta}}{\sigma})-\exp(\frac{-\zeta_{j^*k}\tilde{\beta}}{\sigma})}{2}\right| \leq \frac{|\zeta_{j^*k}|\tilde{\beta}}{\sigma}\frac{\exp(\frac{\zeta_{j^*k}\tilde{\beta}}{\sigma})+\exp(-\frac{\zeta_{j^*k}\tilde{\beta}}{\sigma})}{2},$$

and

$$I \leq \frac{\tilde{\beta}^2}{\sigma}\left(\sum_{k=0}^{2^{j^*}-1}|\delta\beta_{j^*k}||\zeta_{j^*k}|\right) E_\xi Z_\xi \leq \frac{\tilde{\beta}^2}{\sigma}||\delta\beta||_2\left(\sum_{k=0}^{2^{j^*}-1}\zeta^2_{j^*k}\right)^{1/2} E_\xi Z_\xi.$$

??? Due to (28), the right hand side of the latter inequality can be bounded on Γ by

$$32\sqrt{2}\mathrm{e}^{3/2}2^{j^*/2}\tilde{\beta}^2(E_{\beta^0}||\delta\beta||_2^2)^{1/2}\sigma^{-1}. \qquad (33)$$

When substituting into (33) the values of $\tilde{\beta}$ and 2^{j^*} from (27) we obtain the result of the lemma. ∎

Lemma 3 *We have*

$$1) \qquad E_{\beta^0}E_\xi Z_\xi = 1,$$

and for N large enough,

$$2) \qquad E_{\beta^0}[E_\xi Z_\xi]^2 \leq e.$$
$$3) \qquad P_{\beta^0}(A) = P_{\beta^0}(\frac{1}{2} \leq E_\xi Z_\xi \leq 4e) \geq \frac{1}{16e}. \qquad (34)$$

Proof: The proof of 1) is straightforward. To show 2) we decompose

$$E_{\beta^0}[E\xi Z_\xi]^2 = \prod_{k=0}^{2^{j^*}-1} I_k,$$

where

$$I_k = E_{\beta^0}\left[\frac{1}{4}\exp\left(\frac{2\sigma\zeta_{j^*k}\tilde{\beta}-\tilde{\beta}^2}{\sigma^2}\right)+\frac{1}{4}\exp\left(\frac{-2\sigma\zeta_{j^*k}\tilde{\beta}-\tilde{\beta}^2}{\sigma^2}\right)+\frac{1}{2}\exp(\frac{-\tilde{\beta}^2}{\sigma^2})\right].$$

When taking the expectation, we obtain

$$I_k = \frac{1}{2}[\,\mathrm{e}^{\frac{\tilde{\beta}^2}{\sigma^2}} + \mathrm{e}^{-\frac{\tilde{\beta}^2}{\sigma^2}}] < 1 + \frac{\tilde{\beta}^4}{2\sigma^4} + \frac{\tilde{\beta}^8}{12\sigma^8},$$

since $\frac{\tilde{\beta}}{\sigma} < 1$ and $(\frac{\mathrm{e}^x+\mathrm{e}^{-x}}{2})^{(IV)} < 2$ for $0 \leq x \leq 1$.

On the other hand, by the choice of $\tilde{\beta}$ $\frac{\tilde{\beta}^8}{\sigma^8} = o(\frac{\tilde{\beta}^4}{\sigma^4})$, and

$$E_{\beta^0}[E_\xi Z_\xi]^2 \leq \left(1 + 2^{j^*}\frac{\tilde{\beta}^4}{2\sigma^4}(1+o(1))\right) \leq \;\mathrm{e}$$

for N large enough.

3) Let a positive random variable x satisfy $Ex=1$, $Ex^2<\infty$. Then for any $R>1$

$$\int_0^{REx^2} x\mu(dx) \geq 1-1/R,$$

and

$$\int_{1/2}^{REx^2} x\mu(dx) \geq 1-\frac{1}{R}-\int_0^{1/2} x\mu(dx) \geq \frac{1}{2}-\frac{1}{R}.$$

We conclude that

$$\mu(\frac{1}{2} \leq x \leq REx^2) \geq \frac{\frac{1}{2}-\frac{1}{R}}{REx^2}.$$

We finally get for $R=4$

$$\mu(\frac{1}{2} \leq x \leq 4Ex^2) \geq \frac{1}{16Ex^2}.$$

When substituting $E_\xi Z_\xi$ for x we obtain

$$P_\beta(\frac{1}{2} \leq E_\xi Z_\xi \leq 4Ex^2) \geq \frac{1}{16\mathrm{e}}.$$

∎

When summing the results of Lemma 1 and Lemma 2 we get for some C, C', $C''>0$

$$\begin{aligned} r_{\beta^0}(\delta,N) &\geq C\frac{E_{\beta^0}E_\xi\left[Z_\xi(\rho_N^4-4\rho_N^2\tilde\beta\xi^T\delta\beta)1\{\Gamma\}\right]}{(\rho_N+(E\|\delta\beta\|_2^2)^{1/2})^2} \geq C\frac{\rho_N^2 E_{\beta^0}E_\xi[(Z_\xi\rho_N^2-4\tilde\beta Z_\xi\xi^T\delta\beta)1\{\Gamma\}]}{(\rho_N+(E\|\delta\beta\|_2^2)^{1/2})^2} \\ &\geq C'\rho_N^2\left(\frac{\rho_N^2-C''\sigma_e N^{-1/2}(E\|\delta\beta\|_2^2)^{1/2}}{\rho_N+(E\|\delta\beta\|_2^2)^{1/2}}\right)^2 P_{\beta_0}(\Gamma). \end{aligned}$$

Now we are done because by the Tchebychev inequality

$$P_{\beta^0}\left(\|\hat\beta-\beta^0\|_2 \geq 8\sqrt{\mathrm{e}}(E_{\beta^0}\|\hat\beta-\beta^0\|_2^2)^{1/2}\right) \leq \frac{1}{64\mathrm{e}}$$

and

$$\begin{aligned} P_{\beta^0}(\Gamma) &\geq P_{\beta^0}(A)-P_{\beta^0}(B^c)-P_{\beta^0}(C^c) \\ &\geq \frac{1}{16e}-\frac{1}{64\mathrm{e}}-P\left(2^{-j^*}\sum_{k=0}^{2^{j^*}-1}\zeta_{j^*k}^2>2\right) \\ &\geq C>0 \end{aligned}$$

for N large enough. ∎

6.3 Translation into sequence space

We start with the translation of our estimation problem into the space of the sequences of wavelet coefficients. For the sake of simplicity we suppose that $N = 2^{j_0}$. For the computation of wavelet coefficient in the case $N \neq 2^{j_0}$ the reader can refer to [4].

Now

$$f_{j_0}(x) = \alpha' \phi(x) + \sum_{j=0}^{j_0} \sum_k \beta'_{jk} \psi_{jk}(x), \tag{35}$$

where

$$\alpha' = \frac{1}{N} \sum_{i=1}^{N} f(\frac{i}{N}) \phi(\frac{i}{N}), \quad \beta'_{jk} = \frac{1}{N} \sum_{i=1}^{N} f(\frac{i}{N}) \psi_{jk}(\frac{i}{N}).$$

Then the empirical wavelet coefficients satisfy:

$$\widehat{\alpha} = \alpha' + \zeta, \quad y_{jk} = \beta'_{jk} + \zeta_{jk},$$

with

$$\zeta = \frac{1}{N} \sum_{i=1}^{N} w_i \phi(\frac{i}{N}), \quad \zeta_{jk} = \frac{1}{N} \sum_{i=1}^{N} w_i \psi_{jk}(\frac{i}{N}).$$

We present here a summary of properties of the sequence of empirical wavelet coefficients. The next lemma is an immediate corollary of Proposition 2 in [4].

Lemma 4 *Suppose that $f \in \mathcal{F}(s, L)$, $s > 0$. Then there is a constant C_0 (which depends on the wavelet used) such that the sequence $\beta' = (\alpha', \beta'_{jk})$ satisfies*

$$\beta' \in \mathcal{F}(s, C_0 L) \;\; and \;\; \|f - f_{j_0}\|_2 = O(L 2^{-j_0 s}) = O\left(L N^{-s}\right). \tag{36}$$

If we denote

$$m'_N = \|\widehat{f}_N - f_{j_0}\|_2 = \left[\sum_{j=0}^{j_0} \|\widehat{\beta}_{j\cdot} - \beta'_{j\cdot}\|_2^2 \right]^{1/2}$$

then due to (36) we can bound $m_N = \|\widehat{f}_N - f_{j_0}\|_2$ as follows:

$$|m_N - m'_N| \leq \|f - f_{j_0}\|_2 \leq C L N^{-s}. \tag{37}$$

This implies immediately that if $m'_N = \|\widehat{f}_N - f_{j_0}\|_2$, then

$$\left| [E_f(\widehat{m}_N - m_N)^2]^{1/2} - [E_f(\widehat{m}_N - m'_N)^2]^{1/2} \right| \leq C L N^{-s}. \tag{38}$$

So to show the bounds of Theorems 2 and 3 it suffices to control the value $[E_f(\widehat{m}_N - m'_N)^2]^{1/2}$. Furthermore, it follows from (36) that the coefficients β'_{jk} satisfy (up to an "absolute constant") the same norm relation (6) as true coefficients β_{jk}. Since this is the only property of wavelet coefficients used in the study of the estimate $\widehat{m}_N$, with some abuse of notations we substitute in the sequel β'_{jk} for β_{jk}. This gives the model

$$y_{jk} = \beta_{jk} + \zeta_{jk} \tag{39}$$

for empirical wavelet coefficients.

Now note random variables ζ and ζ_{jk} have Gaussian distribution with $E\zeta = E\zeta_{jk} = 0$. Furthermore, since the sequences $\psi_{jk}(\frac{i}{N})$, $i = 1, ..., N$ are orthonormal for different j and k, the variables ζ_{jk} are mutually independent and $E\zeta_{jk}^2 = \frac{\sigma_w^2}{N}$.

6.4 Technical lemmas

Now we establish some technical results for the latter use:

Lemma 5

$$\begin{aligned}
1_{\rho_j \geq 2^j\sigma^2} &\leq 1_{\|\beta_{j\cdot}\|_2^2 \geq \sigma^2 2^{j-3}} + 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 > \sigma^2 2^{j-3}}; && (40)\\
1_{\rho_j < 2^j\sigma^2} &\leq 1_{\|\beta_{j\cdot}\|_2^2 < 9\sigma^2 2^j} + 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 > \sigma^2 2^j}; && (41)\\
1_{\rho_j < \lambda 2^{j/2}\sigma^2} &\leq 1_{\|\beta_{j\cdot}\|_2^2 \leq \lambda 2^{j/2+2}\sigma^2} + 1_{\beta_{j\cdot}^T\zeta_{j\cdot} < -5\sqrt{\lambda}2^{j/4-3}\sigma\|\beta_{j\cdot}\|_2} + 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 < -\lambda 2^{j/2-1}\sigma^2}; && (42)\\
1_{\rho_j \geq \lambda 2^{j/2}\sigma^2} &\leq 1_{\|\beta_{j\cdot}\|_2^2 \geq \lambda 2^{j/2-1}\sigma^2} + 1_{\beta_{j\cdot}^T\zeta_{j\cdot} > \sqrt{\lambda}2^{j/4-7/2}\sigma\|\beta_{j\cdot}\|_2} + 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 > \lambda 2^{j/2-2}\sigma^2}. && (43)
\end{aligned}$$

Furthermore, if $2^{j/2-2} > \lambda$,

$$1_{\lambda 2^{j/2}\sigma^2 \leq \rho_j < 2^j\sigma^2} \leq 1_{\lambda 2^{j/2-1}\sigma^2 \leq \|\beta_{j\cdot}\|_2^2 \leq 2^{j+2}\sigma^2} + 1_{|\beta_{j\cdot}^T\zeta_{j\cdot}| > \lambda\sigma\|\beta_{j\cdot}\|_2} + 1_{|\,\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2| > \lambda 2^{j/2-2}\sigma^2}; \tag{44}$$

and for $2^{j/2-2} \leq \lambda$, $\lambda \geq 5/2$*:*

$$1_{\rho_j < 2^j\sigma^2} \leq 1_{\|\beta_{j\cdot}\|_2^2 \leq 48\lambda^3\sigma^2} + 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 > \lambda 2^j\sigma^2} \tag{45}$$

Proof: Let us show (40). Note that

$$4\|\beta_{j\cdot}\|_2^2 + \frac{4}{3}\|\zeta_{j\cdot}\|_2^2 \geq \|\beta_{j\cdot} + \zeta_{j\cdot}\|_2^2,$$

so that

$$1_{\rho_j \geq 2^j\sigma^2} \quad \leq \quad 1_{4\|\beta_{j\cdot}\|_2^2 + \frac{4}{3}\|\zeta_{j\cdot}\|_2^2 \geq 2^{j+1}\sigma^2} \leq 1_{4\|\beta_{j\cdot}\|_2^2 \geq 2^j\sigma^2} + 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 > \sigma^2 2^{j-3}}.$$

The proof of (41) is conducted in the same way. Let us consider (42):

$$
\begin{aligned}
& 1_{\|y_{j\cdot}\|_2^2-2^j\sigma^2<\lambda 2^{j/2}\sigma^2} = 1_{\|\beta_{j\cdot}\|_2^2+2\beta_{j\cdot}^T\zeta_{j\cdot}+\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2<\lambda 2^{j/2}\sigma^2} \\
& \quad \le \quad 1_{\|\beta_{j\cdot}\|_2^2+2\beta_{j\cdot}^T\zeta_{j\cdot}\le\frac{3\lambda}{2}2^{j/2}\sigma^2} + 1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2<-\lambda 2^{j/2-1}\sigma^2} \\
& \quad \le \quad 1_{\|\beta_{j\cdot}\|_2^2\le\lambda 2^{j/2+2}\sigma^2} + 1_{\|\beta_{j\cdot}\|_2^2>\lambda 2^{j/2+2}\sigma^2} 1_{2\beta_{j\cdot}^T\zeta_{j\cdot}<-\|\beta_{j\cdot}\|_2^2+\frac{3\lambda}{2}2^{j/2}\sigma^2} + 1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2<-\lambda 2^{j/2-1}\sigma^2} \\
& \quad \le \quad 1_{\|\beta_{j\cdot}\|_2^2\le\lambda 2^{j/2+2}\sigma^2} + 1_{\beta_{j\cdot}^T\zeta_{j\cdot}<-5\sqrt{\lambda}2^{j/4-3}\|\beta_{j\cdot}\|_2\sigma} + 1_{|\zeta_{j\cdot}\|_2^2-2^j\sigma^2<-\lambda 2^{j/2-1}\sigma^2}.
\end{aligned}
$$

The proof of (43) and (44) follows the same lines. To show (45) we note that

$$\rho_j + 2^j\sigma^2 = \|\beta_{j\cdot} + \zeta_{j\cdot}\|_2^2 \ge \frac{1}{3}\|\beta_{j\cdot}\|_2^2 - \frac{1}{2}\|\zeta_{j\cdot}\|_2^2.$$

Thus for any $\gamma > \frac{5}{2}\sigma^2 2^j$

$$1_{\rho_j<2^j\sigma^2} \quad \le \quad 1_{\frac{1}{3}\|\beta_{j\cdot}\|_2^2-\frac{1}{2}\|\zeta_{j\cdot}\|_2^2<2\sigma^2 2^j} \le 1_{\|\beta_{j\cdot}\|_2^2\le 3\gamma} + 1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2>2\gamma-5\sigma^2 2^j}.$$

Now the choice $\gamma = \frac{\lambda+5}{2}2^j\sigma^2 \le \lambda 2^j\sigma^2$ in the last formula gives

$$
\begin{aligned}
1_{\rho_j<2^j\sigma^2} \quad & \le \quad 1_{\|\beta_{j\cdot}\|_2^2\le 3\lambda 2^j\sigma^2} + 1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2>\lambda 2^j\sigma^2} \\
& \le \quad 1_{\|\beta_{j\cdot}\|_2^2\le 48\lambda^3\sigma^2} + 1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2>\lambda 2^j\sigma^2}.
\end{aligned}
$$

∎

Lemma 6 *Let $(\zeta_i) \in \mathbf{R}^n$ be a vector of Gaussian independent and identically distributed random variables, $E\zeta_1 = 0$, $E\zeta_1^2 = \sigma^2$. Then for $0 < h \le n^{1/6}$*

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n(\zeta_i^2 - \sigma^2) > h\sigma^2\right) < e^{\frac{-(h^2-1)}{4}}.$$

Proof: First note that

$$E\exp(\alpha(\zeta_1^2 - \sigma^2)) = \frac{e^{-\alpha\sigma^2}}{(1-2\alpha\sigma^2)^{1/2}} = \exp(-\alpha\sigma^2 - \frac{1}{2}\ln(1-2\alpha\sigma^2)). \tag{46}$$

On the other hand, for $\alpha\sigma^2$ small enough, the logarithmic term can be written as

$$\frac{1}{2}\ln(1-2\alpha\sigma^2) = -\alpha\sigma^2 - \alpha^2\sigma^4\left(1 + \sum_{k=1}^\infty \frac{2^{k+1}(\alpha\sigma^2)^k}{k+2}\right) \ge -\alpha\sigma^2 - \alpha^2\sigma^4(1+2\alpha\sigma^2) \tag{47}$$

for $\alpha\sigma^2 \le 1/6$. Due to (46) and (47) we have by the Tchebychev inequality

$$\begin{aligned}
P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\zeta_i^2-\sigma^2) > h\sigma^2\right) &\le E\exp\left(\alpha\sum_{i=1}^{n}(\zeta_i^2-\sigma^2)-\alpha h\sigma^2\sqrt{n}\right)\\
&= E\exp\left(n\alpha(\zeta_1^2-\sigma^2)-\alpha h\sigma^2\sqrt{n}\right)\\
&\le \exp(n\alpha^2\sigma^4(1+2\alpha\sigma^2)-\alpha h\sigma^2\sqrt{n}).
\end{aligned}$$

If we take $\alpha = \frac{h}{2\sigma^2\sqrt{n}}$, we obtain

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\zeta_i^2-\sigma^2) > h\sigma^2\right) \le \mathrm{e}^{-\frac{h^2}{4}+\frac{h^3}{4\sqrt{n}}},$$

what gives the lemma. ∎

Lemma 7 *Let $\beta_j \in \mathbf{R}^{2^j}$ and $\zeta_j \in \mathbf{R}^{2^j}$ be a Gaussian random vector with zero mean and the covariance matrix $E\zeta_j\zeta_j^T = \sigma I$. Then there is $C<\infty$*

$$\left[E\left(\beta_{j\cdot}^T\zeta_{j\cdot}\right)^2\right]^{1/2} = \sigma\|\beta_{j\cdot}\|_2,\quad \left[E\left(\beta_{j\cdot}^T\zeta_{j\cdot}\right)^4\right]^{1/4} \le C2^{j/4}\sigma\|\beta_{j\cdot}\|_2,\quad \left[E(\|\zeta_{j\cdot}\|^2-2^j\sigma^2)^4\right]^{1/4} \le C2^{j/2}\sigma^2$$

for some $C<\infty$;

$$P\left(|\beta_{j\cdot}^T\zeta_{j\cdot}| > \sqrt{\lambda}2^{j/4-7/2}\sigma\|\beta_{j\cdot}\|_2\right) \le 2\exp(-\frac{\lambda^2}{64}),\quad \textit{for } 2^{j/2-2}\ge\lambda; \tag{48}$$

and for $0<\lambda<2^{j/6}$

$$P\left(\left|\,\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2\right| > \lambda 2^{j/2-2}\sigma^2\right) \le 4\exp(-\frac{\lambda^2}{64}). \tag{49}$$

Proof: Recall that $\beta_{j\cdot}^T\zeta_{j\cdot}$ is a Gaussian random variable with $E\beta_{j\cdot}^T\zeta_{j\cdot} = 0$. Thus

$$E(\beta_{j\cdot}^T\zeta_{j\cdot})^2 = \sigma^2\|\beta_{j\cdot}\|_2^2 \tag{50}$$

and

$$E(\beta_{j\cdot}^T\zeta_{j\cdot})^4 \le \|\beta\|_4^4 E\|\zeta_{j\cdot}\|_4^4 \le 3\|\beta\|_2^4 2^j\sigma^4.$$

In the same way

$$E(\|\zeta_{j\cdot}\|^2-2^j\sigma^2)^4 \le C2^{2j}\sigma^8.$$

Furthermore, (50) implies that

$$P\left(|\beta_{j\cdot}^T\zeta_{j\cdot}| > 5\sqrt{\lambda}2^{j/4-7/2}\sigma\|\beta_{j\cdot}\|_2\right) \le 2\exp(-\frac{\lambda 2^{j/2-2}}{64}) \le 2\exp(-\frac{\lambda^2}{64}).$$

The bound (49) follows immediately from Lemma 6. ∎

6.5 Proof of Theorem 2

Due to (39) and by construction, the estimator $\widehat{\beta}_N$ in (16), the error m'_N satisfies:

$$(m'_N)^2 = \sum_{j=0}^{j_0} \|\widehat{\beta}_{j\cdot} - \beta_{j\cdot}\|_2^2 = \sum_{j=0}^{j_0} \left(\|\zeta_{j\cdot}\|_2^2 1_{\rho_j \geq 2^j\sigma^2} + \|\beta_{j\cdot}\|_2^2 1_{\rho_j < 2^j\sigma^2} \right).$$

Then we have for the difference $\widehat{m}_N^2 - (m'_N)^2$:

$$\begin{aligned} |\widehat{m}_N^2 - (m'_N)^2| &\leq \left| -\sum_{j=0}^{j_0} (\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2) 1_{\rho_j \geq 2^j\sigma^2} + \sum_{j=0}^{j^*} (\rho_j - \|\beta_{j\cdot}\|_2^2) 1_{\rho_j < 2^j\sigma^2} - \sum_{j=j^*+1}^{j_0} \|\beta_{j\cdot}\|_2^2 \right| \\ &\leq \left| \sum_{j=0}^{j^*} (\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2) \right| + \left| \sum_{j=j^*+1}^{j_0} (\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2) 1_{\rho_j \geq 2^j\sigma^2} \right| \\ &\quad +2 \left| \sum_{j=0}^{j^*} \beta_{j\cdot}^T \zeta_{j\cdot} 1_{\rho_j < 2^j\sigma^2} \right| + \sum_{j=j^*+1}^{\infty} \|\beta_{j\cdot}\|_2^2 = \sum_{i=1}^{4} \delta_N^{(i)}. \end{aligned} \tag{51}$$

The following two estimates are immediate:

$$\delta_N^{(4)} \leq CL^2 2^{-2j^*s^*} \quad \text{and} \quad \left[E(\delta_N^{(1)})^2\right]^{1/2} \leq C 2^{j^*/2}\sigma^2. \tag{52}$$

Lemma 8 $\left[E(\delta_N^{(2)})^2\right]^{1/2} \leq C\sigma^2\sqrt{N}\exp(-\frac{9\,2^{j^*}}{16})$.

Proof: We use (40) to obtain

$$\begin{aligned} \left[E(\delta_N^{(2)})^2\right]^{1/2} &\leq \left[E\left(\sum_{j=j^*+1}^{j_0} (\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2) 1_{\|\beta_{j\cdot}\|_2^2 \geq 2^{j-3}\sigma^2} \right)^2 \right]^{1/2} \\ &\quad + \left[E\left(\sum_{j=j^*+1}^{j_0} (\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2) 1_{\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 > 2^{j-3}\sigma^2} \right)^2 \right]^{1/2} \\ &\leq C \sum_{j=j^*+1}^{j_0} 2^{j/2}\sigma^2 P^{1/2}\left(\left| \|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 \right| > 3\,2^{j-1}\sigma^2 \right) \leq C'\sqrt{N}\sigma^2 \exp(-\frac{9\,2^{j^*}}{16}). \end{aligned}$$

∎

Lemma 9 $\left[E(\delta_N^{(3)})^2\right]^{1/2} \leq C\left(2^{j^*/2}\sigma^2 + \sigma^2 \log N\right)$.

Proof: Let $\lambda = 16\sqrt{\log N}$ and j_1 satisfy $4\lambda \le 2^{j_1/2} < 8\lambda$. Due to (41) and (45) we have

$$\begin{aligned}
\left[E(\delta_N^{(3)})^2\right]^{1/2} &\le 2\left[E\left(\sum_{j=0}^{j_1-1}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\rho_j<2^j\sigma^2}\right)^2\right]^{1/2} + 2\left[E\left(\sum_{j=j_1}^{j^*}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\rho_j<2^j\sigma^2}\right)^2\right]^{1/2}\\
&\le 2\left[E\left(\sum_{j=0}^{j_1-1}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\|\beta_{j\cdot}\|_2^2<64\lambda^3\sigma^2}\right)^2\right]^{1/2} + 2\left[E\left(\sum_{j=0}^{j_1-1}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2>\lambda 2^j\sigma^2}\right)^2\right]^{1/2}\\
&\quad +2\left[E\left(\sum_{j=j_1}^{j^*}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\|\beta_{j\cdot}\|_2^2<9\sigma^2 2^j}\right)^2\right]^{1/2} + 2\left[E\left(\sum_{j=j_1}^{j^*}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2>\sigma^2 2^j}\right)^2\right]^{1/2}.
\end{aligned}$$

Now Lemma 7 supplies the bound (cf. the proof of Lemma 11):

$$\begin{aligned}
\left[E(\delta_N^{(3)})^2\right]^{1/2} &\le C\left(\lambda^2\sigma^2 + \lambda L\sigma\exp(-\frac{\lambda^2}{256}) + 2^{j^*/2}\sigma^2 + L\sigma\sum_{j=j_1}^{j^*}2^{j/4}\exp(-\frac{2^{2j}-1}{4})\right)\\
&\le C\left(2^{j^*/2}\sigma^2 + \sigma^2\log N + \frac{L\sigma\sqrt{\log N}}{N}\right).
\end{aligned}$$

∎

We now substitute the bound of (52) and those in Lemmas 8 and 9 into (51) to obtain

$$\left[E(\widehat{m}_N^2 - (m')_N^2)^2\right]^{1/2} \le C\left(2^{j^*/2}\sigma^2 + L^2 2^{-2j^*s^*} + \sigma^2\log N\right) \le C(L^*)^{2/(4s^*+1)}\sigma^{8s^*/(4s^*+1)}. \quad (53)$$

Note that for any $\gamma > 0$

$$\begin{aligned}
(\widehat{m}_N - m_N)^2 &= (\widehat{m}_N - m_N)^2 1_{m_N\ge\gamma} + (\widehat{m}_N - m_N)^2 1_{m_N<\gamma}\\
&\le \frac{(\widehat{m}_N^2 - m_N^2)^2}{\gamma^2} + 2\widehat{m}_N^2 + 2\gamma^2\\
&\le \frac{(\widehat{m}_N^2 - m_N^2)^2}{\gamma^2} + 2(\widehat{m}_N^2 - m_N^2) + 4\gamma^2. \qquad (54)
\end{aligned}$$

We set

$$\gamma = \sqrt{(L^*)^{2/(4s^*+1)}\sigma^{8s^*/(4s^*+1)} + \sigma^2\log N}.$$

Then (20) follows from (38) and (53). ∎

6.6 Proof of Theorem 3

Recall that it holds for the error m'_N of the estimator $\widehat{\beta}_N$:

$$(m'_N)^2 = \sum_{j=0}^{j_0} \left(\|\zeta_{j\cdot}\|_2^2 1_{\rho_j \geq 2^j\sigma^2} + \|\beta_{j\cdot}\|_2^2 1_{\rho_j < 2^j\sigma^2} \right).$$

Consider the following decomposition of the set $\{j:\ j \leq j_0\}$:

$$\begin{aligned} J_0 &= \{j \leq j_0:\ \rho_j \geq \sigma^2 2^j\}; && (55)\\ J_1 &= \{j \leq j_0:\ \lambda\sigma^2 2^{j/2} \leq \rho_j < \sigma^2 2^j\}; && (56)\\ J_2 &= \{j \leq j_0\}/\{J_0 \cup J_1\}. && (57) \end{aligned}$$

Then the error $\widehat{m}_N - m_N$ can be represented as follows:

$$\begin{aligned} \widehat{m}_N^2 - (m'_N)^2 &= \sum_{j\in J_0} \|\zeta_{j\cdot}\|^2 - 2^j\sigma^2 + \sum_{j\in J_1} (2\beta_{j\cdot}^T\zeta_{j\cdot} + \|\zeta_{j\cdot}\|^2 - 2^j\sigma^2)\\ &\quad + \sum_{j\in J_2} \|\beta_{j\cdot}\|_2^2 + \sum_{j>j_0} \|\beta_{j\cdot}\|_2^2\\ &= \sum_{j\in J_0\cup J_1} \|\zeta_{j\cdot}\|^2 - 2^j\sigma^2 + \sum_{j\in J_1} 2\beta_{j\cdot}^T\zeta_{j\cdot} + \sum_{j\in J_2} \|\beta_{j\cdot}\|_2^2\\ &= \sum_{i=1}^{3} \delta_N^{(i)}. \qquad (58) \end{aligned}$$

Let j_1 be such that

$$\{4\lambda < 2^{j_1/2} \leq 8\lambda\}. \qquad (59)$$

We define also

$$j_2 = \max\{j_1 < j \leq j_0:\ 2\|\beta_{j\cdot}\|_2^2 \geq \lambda 2^{j/2}\sigma^2\} \qquad (60)$$

Clearly, for N large enough, $0 < j_1 \leq j_2 \leq j_0$.

Lemma 10 $\left[E(\delta_N^{(1)})^2\right]^{1/2} \leq C'\left(2^{j_2/2}\sigma^2 + \sigma^2\sqrt{N}\exp(-\frac{\lambda^2}{256})\right).$

Proof: By the Minkowski inequality we get from (43):

$$\left[E(\delta_N^{(1)})^2\right]^{1/2} = E\left[\left(\sum_{j=0}^{j_0} \|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2\right)^2 1_{\rho_j \geq \lambda\sigma^2 2^{j/2}}\right]^{1/2}$$

$$
\begin{aligned}
\leq \quad & \left[E\left(\sum_{j=0}^{j_1-1}||\zeta_{j\cdot}||_2^2-2^j\sigma^2\right)^2\right]^{1/2}+\left[E\left(\sum_{j=j_1}^{j_0}||\zeta_{j\cdot}||_2^2-2^j\sigma^2\right)^2 1_{||\beta_{j\cdot}||^2\geq\lambda\sigma^2 2^{j/2-1}}\right]^{1/2} \\
& +\left[E\left(\sum_{j=j_1}^{j_0}||\zeta_{j\cdot}||^2-2^j\sigma^2\right)^2 1_{\beta_{j\cdot}^T\zeta_{j\cdot}\geq\sqrt{\lambda}2^{j/4-5/2}||\beta_{j\cdot}||_2\sigma}\right]^{1/2} \\
& +\left[E\left(\sum_{j=j_1}^{j_0}||\zeta_{j\cdot}||^2-2^j\sigma^2\right)^2 1_{||\zeta_{j\cdot}||^2-2^j\sigma^2\geq\lambda 2^{j/2-2}\sigma^2}\right]^{1/2}.
\end{aligned}
$$

Using the Minkowski inequality again and the definition of j_2 above we obtain:

$$
\begin{aligned}
\left[E(\delta_N^{(1)})^2\right]^{1/2} \quad \leq \quad & C2^{j_1}\sigma^2+\sum_{j=j_1}^{j_2}\left[E(||\zeta_{j\cdot}||^2-2^j\sigma^2)^2\right]^{1/2} \\
& +\sum_{j=j_1}^{j_0}\left[E(||\zeta_{j\cdot}||^2-2^j\sigma^2)^4\right]^{1/4}P^{1/4}\left(\beta_{j\cdot}^T\zeta_{j\cdot}\geq\sqrt{\lambda}2^{j/4-5/2}||\beta_{j\cdot}||_2\sigma\right) \\
& +\sum_{j=0}^{j_0}\left[E(||\zeta_{j\cdot}||^2-2^j\sigma^2)^4\right]^{1/4}P^{1/4}\left(||\zeta_{j\cdot}||^2-2^j\sigma^2\geq\lambda 2^{j/2-2}\sigma^2\right)
\end{aligned}
$$

Due to the bounds in Lemma 7 we conclude that

$$
\begin{aligned}
\left[E(\delta_N^{(1)})^2\right]^{1/2} \quad & \leq \quad C\left(\sum_{j=0}^{j_2}2^{j/2}\sigma^2+\sum_{j=0}^{j_0}2^{j/2}\sigma^2\exp(-\frac{\lambda^2}{256})\right) \\
& \leq \quad C'\left(2^{j_2/2}\sigma^2+\sigma^2\sqrt{N}\exp(-\frac{\lambda^2}{256})\right).
\end{aligned}
$$

∎

Lemma 11 $\left[E(\delta_N^{(2)})^2\right]^{1/2}\leq C\left(2^{j_2/2}\sigma^2+\lambda^2\sigma^2+LN^{1/4}(\log N)\sigma\exp(-\frac{\lambda^2}{256})\right)$.

Proof: Using (44) and (45) we decompose $\delta_N^{(2)}$ as follows:

$$
\left[E(\delta_N^{(2)})^2\right]^{1/2} \quad = \quad \left[E\left(2\sum_{j=0}^{j_0}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\lambda 2^{j/2}\sigma^2\leq\rho_j\leq 2^j\sigma^2}\right)^2\right]^{1/2}
$$

$$
\begin{aligned}
& \leq && 2\left[E\left(\sum_{j=0}^{j_1-1}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\lambda 2^{j/2}\sigma^2\leq\rho_j\leq 2^j\sigma^2}\right)^2\right]^{1/2}+2\left[E\left(\sum_{j=j_1}^{j_0}\beta_{j\cdot}^T\zeta_{j\cdot}1_{\lambda 2^{j/2}\sigma^2\leq\rho_j\leq 2^j\sigma^2}\right)^2\right]^{1/2} \\
\text{(due to (45))} & \leq && 2\left(\sum_{j=0}^{j_1-1}\|\beta_{j\cdot}\|_2^2\sigma^2 1_{\|\beta_{j\cdot}\|_2^2\leq 48\lambda^3\sigma^2}\right)^{1/2} \\
& && +2\sum_{j=0}^{j_1-1}\left[E(\beta_{j\cdot}^T\zeta_{j\cdot})^4\right]^{1/4}P^{1/4}\left(\|\zeta_{j\cdot}\|_2-2^j\sigma^2>\lambda 2^j\sigma^2\right) \\
\text{(by (44))} & && +2\sum_{j=j_1}^{j_0}\|\beta_{j\cdot}\|_2\sigma 1_{\lambda 2^{j/2-1}\sigma^2\leq\|\beta_{j\cdot}\|_2^2\leq 2^{j+2}\sigma^2} \\
& && +2\sum_{j=j_1}^{j_0}\left[E(\beta_{j\cdot}^T\zeta_{j\cdot})^4\right]^{1/4}P^{1/4}\left(|\beta_{j\cdot}^T\zeta_{j\cdot}|>\sqrt{\lambda}2^{j/4-7/2}\sigma\|\beta_{j\cdot}\|_2\right) \\
& && +2\sum_{j=j_1}^{j_0}\left[E(\beta_{j\cdot}^T\zeta_{j\cdot})^4\right]^{1/4}P^{1/4}\left(|\,\|\zeta_{j\cdot}\|_2-2^j\sigma^2|>\lambda 2^{j/2-2}\sigma^2\right)
\end{aligned}
$$

Then Lemma 7 gives

$$
\begin{aligned}
\left[E(\delta_N^{(2)})^2\right]^{1/2} & \leq C\left(\lambda^2\sigma^2+2^{j_1/4}\sigma\|\beta\|_2\exp(-\frac{\lambda^2}{256})\right. \\
& \left.+\sum_{j=j_1}^{j_2}2^{j/2}\sigma^2+\|\beta\|_2\sum_{j=j_1}^{j_0}2^{j/4}\sigma\exp(-\frac{\lambda^2}{256})+\|\beta\|_2\sum_{j=j_1}^{j_0}2^{j/2}\sigma\exp(-\frac{\lambda^2}{256})\right) \\
& \leq C'\left(2^{j_2/2}\sigma^2+\lambda^2\sigma^2+LN^{1/4}(\log N)\sigma\exp(-\frac{\lambda^2}{256})\right).
\end{aligned}
$$

∎

Lemma 12 $\left[E(\delta_N^{(3)})^2\right]^{1/2}\leq C'\left(\lambda^4\sigma^2+L^2\exp(-\frac{\lambda^2}{64})+\lambda 2^{j/3}\sigma^2+L^2 2^{-2j_2 s}\right)$.

Proof: We decompose $\delta_N^{(3)}$ using (42) and (45):

$$
\begin{aligned}
\left[E(\delta_N^{(3)})^2\right]^{1/2} & = \left[E\left(\sum_{j=0}^{j_0}\|\beta_{j\cdot}\|_2^2 1_{\rho_j<\lambda 2^{j/2}\sigma^2}\right)^2\right]^{1/2} \qquad (61)\\
& \leq \sum_{j=0}^{j_1-1}\|\beta_{j\cdot}\|_2^2 1_{\|\beta_{j\cdot}\|_2^2\leq 48\lambda^3\sigma^2}+\sum_{j=0}^{j_1-1}\|\beta_{j\cdot}\|_2^2 P^{1/2}\left(\|\zeta_{j\cdot}\|_2^2-2^j\sigma^2>\lambda 2^j\sigma^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{j=j_1}^{j_0} \|\beta_{j\cdot}\|_2^2 P^{1/2}\left(\beta_{j\cdot}^T\zeta_{j\cdot} < -5\sqrt{\lambda}2^{j/4-3}\sigma\|\beta_{j\cdot}\|_2\right)\\
&+\sum_{j=j_1}^{j_0} \|\beta_{j\cdot}\|_2^2 P^{1/2}\left(\|\zeta_{j\cdot}\|_2^2 - 2^j\sigma^2 < -\lambda 2^{j/2-1}\sigma^2\right)\\
&+\left[\sum_{j=j_1}^{j_0} \|\beta_{j\cdot}\|_2^2 1_{\|\beta_{j\cdot}\|_2^2<\lambda 2^{j/2+2}\sigma^2}\right]^{1/2}. \qquad (62)
\end{aligned}
$$

We can decompose the last term of the sum (62) as follows:

$$
\sum_{j=j_1}^{j_0} \|\beta_{j\cdot}\|_2^2 1_{\|\beta_{j\cdot}\|_2^2<\lambda 2^{j/2+2}\sigma^2} \leq \sum_{j=j_1}^{j_2} \lambda 2^{j/2+2}\sigma^2 + \sum_{j=j_2+1}^{j_0} \|\beta_{j\cdot}\|_2^2 \leq C(\lambda 2^{j_2/2}\sigma^2 + L^2 2^{-2j_2 s})
$$

Then we conclude from (62) that

$$
\begin{aligned}
\left[E(\delta_N^{(3)})^2\right]^{1/2} &\leq C\left(\lambda^4\sigma^2 + \|\beta\|_2^2\exp(-\frac{\lambda^2}{4}) + \|\beta\|_2^2\exp(-\frac{\lambda^2}{8})\right.\\
&\quad \left. + \|\beta\|_2^2\exp(-\frac{\lambda^2}{128}) + \lambda 2^{j/3}\sigma^2 + L2^{-2j_2 s}\right)\\
&\leq C'\left(\lambda^4\sigma^2 + L^2\exp(-\frac{\lambda^2}{128}) + \lambda 2^{j/3}\sigma^2 + L^2 2^{-2j_2 s}\right).
\end{aligned}
$$

∎

When summing up the results of Lemmas 10 − 12 we obtain

$$
\left[E(\widehat{m}_N^2 - (m_N')^2)^2\right]^{1/2} \leq C\left(\lambda 2^{j_2/2}\sigma^2 + (\sigma^2 N + L^2)\exp(-\frac{\lambda^2}{256}) + \lambda^4\sigma^2 + L^2 2^{-2j_2 s}\right)
$$

Now, by the definition of the class $\mathcal{F}(s,L)$ we have the bound on 2^{j_2}. Indeed, $\|\beta_{j\cdot}\|_2^2 \leq L^2 2^{-2js}$, thus

$$
\lambda 2^{(j_2+1)/2}\sigma^2 > 2L^2 2^{-2(j_2+1)s}
$$

(cf. the definition (60)). This implies that

$$
2^{j_2} < \left(\frac{\sqrt{2}L^2 2^{-2s}}{\lambda\sigma^2}\right)^{\frac{2}{4s+1}}.
$$

Then the choice $\lambda = 16\sqrt{\log N}$ gives

$$
\left[E(\widehat{m}_N^2 - (m_N')^2)^2\right]^{1/2} \leq C\left(\sigma^2\log^2 N + L^{2/(4s+1)}\left(\sigma^2\sqrt{\ln N}\right)^{4s/(4s+1)}.\right).
$$

Along with (38) this implies that

$$\left[E(\widehat{m}_N^2 - m_N^2)^2\right]^{1/2} \le C\left(\sigma^2 \log^2 N + L^{2/(4s+1)}\left(\sigma^2\sqrt{\ln N}\right)^{4s/(4s+1)} + L^2 N^{-2s}\right). \quad (63)$$

If we take

$$\gamma = \sqrt{\sigma^2 \log^2 N + \frac{L^2}{N} + L^{2/(4s+1)}\left(\sigma^2\sqrt{\ln N}\right)^{4s/(4s+1)} + L^2 N^{-2s}},$$

we obtain from (53)

$$\left[E(\widehat{m}_N - m_N)^2\right]^{1/2} \le C\left(\sigma \log N + L^{1/(4s+1)}\left(\sigma^2\sqrt{\ln N}\right)^{2s/(4s+1)} + LN^{-s}\right).$$

what finishes the proof of the theorem. ∎

Acknowledgements

The authors consider their pleasant duty to thank Sophie Lambert for her pertinent comments on the manuscript.

References

[1] L. Birgé and P. Massart, From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam: research papers in probability and statistics*, ed. D. Pollard, New York, Springer, 55-87, 1997.

[2] A. Cohen, I. Daubechies, B. Jawerth and P. Vial, Multiresolution analysis, wavelets and fast algorithms on an interval, *C.R.* 316, 417-421, 1992.

[3] I. Daubechies, *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.

[4] B. Delyon, A. Juditsky On the Computation of Wavelet Coefficients, *Journal of Approximation Theory*, 88, N. 1, 47-79, 1997

[5] D. Donoho, I. Johnstone, *Minimax Estimation via Wavelet Shrinkage*, Technical Report, Department of Statistics, Stanford University, 1992.

[6] D. Donoho, I. Johnstone, *Adapting to Unknown Smoothness via Wavelet Shrinkage*, Technical Report, Department of Statistics, Stanford University, 1993.

[7] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, *Wavelet Shrinkage: Asymptotia?* Manuscript, 1993.

[8] S. Efromovich and M. Low, On optimal adaptive estimation of a quadratic functional, *Annals of Stat.*, 24, N. 1, 1106-1125, 1996.

[9] S. Efromovich and M. Pinsker, Learning algorithm for nonparametric filtering, *Autom. Remote Control*, **45**, 1434-1440, 1984; translation from *Avtom. Telemekh.*, No.11, 58-65, 1984.

[10] S. Efromovich, Sequential nonparametric estimation with assigned risk, *Ann. Stat.*, **23**, 4, 1376-1392, 1995.

[11] Ibragimov, R. Khasminskij, *Statistical Estimation, Asymptotic Theory.* Moscow, Nauka, 1979 (in Russian). Translation: Springer-Verlag, Berlin, 1981.

[12] A. Juditsky Wavelet Estimators: Adapting to Unknown Smoothness, *Math. Methods of Stat.*, 1, 1-20, 1997.

[13] O. Lepski, Asymptotically minimax adaptive estimation I. *TVIP*, 36, N. 4, 645-659, 1991 (in Russian). Translated as *SIAM J. Theory of Probab. and Appl.*.

[14] M. Low, On nonparametric confidence intervals, *Annals of Stat.*, **25**, 2547-2554, 1997.

[15] Y. Meyer, *Ondelettes* Hermann, Paris, 1990.

[16] H. Triebel, *Theory of Function Spaces II.* Birkhäuser Verlag, Berlin, 1992.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399